

Research Article

Convergence Analysis of an Empirical Eigenfunction-Based Ranking Algorithm with Truncated Sparsity

Min Xu,¹ Qin Fang,² and Shaofan Wang³

¹ School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

² College of Information, Dalian University, Dalian 116622, China

³ Beijing Key Laboratory of Multimedia Intelligent Software Technology, College of Metropolitan Transportation, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Min Xu; wolf_hsu@dlut.edu.cn

Received 13 April 2014; Revised 8 July 2014; Accepted 9 July 2014; Published 21 July 2014

Academic Editor: Sergei V. Pereverzyev

Copyright © 2014 Min Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study an empirical eigenfunction-based algorithm for ranking with a data dependent hypothesis space. The space is spanned by certain empirical eigenfunctions which we select by using a truncated parameter. We establish the representer theorem and convergence analysis of the algorithm. In particular, we show that under a mild condition, the algorithm produces a satisfactory convergence rate as well as sparse representations with respect to the empirical eigenfunctions.

1. Introduction

Motivated by various applications including problems related to information retrieval, user-preference modeling, and computational biology, the problem of ranking has recently gained much attention in machine learning (see, e.g., [1–4]). This paper proposes a kernel-based ranking algorithm to search a ranking function in a data dependent hypothesis space. The space is spanned by certain empirical eigenfunctions which we select by using a truncated parameter. The notion of empirical eigenfunctions, first studied for learning algorithms in [5], has been used to develop classification and regression algorithms which are shown to have strong learning ability [6, 7]. We will use such an idea to develop learning algorithms for ranking.

1.1. The Ranking Problem. The problem of ranking is distinct from both classification and regression. In ranking, one learns a real-valued function that assigns scores to instances, but the scores themselves do not matter; instead, what is important is the relative ranking of instances induced by those scores.

Formally, the problem of ranking may be modeled in the framework of statistical learning theory (see, e.g., [8] for more details). Assume ρ is a Borel probability measure on $Z = X \times Y$, where X is a compact metric space (input or instance

space) and $Y = [0, M]$ (output space) for some $M > 0$. Let ρ_X be its marginal distribution on X and let $\rho(\cdot | x)$ be the conditional distribution on Y at given x . The learner is given a set of samples $\mathbf{z} = \{z_i\}_{i=1}^m = \{(x_i, y_i)\}_{i=1}^m \in Z^m$ drawn independently and identically according to ρ and the goal is to find a function $f_{\mathbf{z}} : X \rightarrow \mathbb{R}$ that ranks future instances with larger labels higher than those with smaller labels. In other words, x is to be ranked as preferred over x' if $f_{\mathbf{z}}(x) > f_{\mathbf{z}}(x')$ and lower than x' if $f_{\mathbf{z}}(x) < f_{\mathbf{z}}(x')$ ($f_{\mathbf{z}}(x) = f_{\mathbf{z}}(x')$ indicates that there is no difference in ranking preference between the two instances). In this setting, the penalty of a ranking function f on a pair of instances (x, x') with corresponding labels y and y' can be taken to be the least squares ranking loss:

$$\mathcal{L}(f, (x, y), (x', y')) = ((y - y') - (f(x) - f(x')))^2, \quad (1)$$

and, as a result, the quality of f can be measured by its expected ranking error:

$$\mathcal{E}(f) = \int_Z \int_Z (y - y' - (f(x) - f(x')))^2 d\rho(x, y) d\rho(x', y'). \quad (2)$$

Let $L_{\rho_X}^2$ be the space of square integrable functions on X with respect to the measure $d\rho_X$. Let \mathcal{G} be the collection of target functions which are defined to be the functions minimizing the error $\mathcal{E}(f)$ over $L_{\rho_X}^2$; that is, $\mathcal{G} = \{f \in L_{\rho_X}^2 : f = \arg \min_{f \in L_{\rho_X}^2} \mathcal{E}(f)\}$. It is apparent from (2) that the error of any ranking function of the form $f + c$ is the same as that of f , where $c \in \mathbb{R}$ is some constant. Therefore, unlike the target function in classification or regression, the target function in ranking is not unique in general. It is easy to show that the regression function of ρ defined by

$$f_\rho(x) = \int_Y y d\rho(y | x), \quad x \in X \quad (3)$$

is a minimizer of the error (2), which indicates that any function of the form $f_\rho + c$ is also a target function. On the other hand, any target function in $L_{\rho_X}^2$ must have the form $f_\rho + c$, which can be checked from Lemma 11 in [9]. Thus we conclude that the collection \mathcal{G} consists exactly of functions that have the form $f_\rho(x) + c$ with $c \in \mathbb{R}$ being an arbitrary constant; that is, $\mathcal{G} = \{f \in L_{\rho_X}^2 : f = f_\rho(x) + c, c \in \mathbb{R}\}$.

1.2. The Mercer Kernel and Empirical Eigenpair. Our algorithm is based on a Mercer kernel and we need to introduce some notions related to kernels (see [10, 11] for more details). Recall that a Mercer kernel is defined to be a symmetric continuous function $K : X \times X \rightarrow \mathbb{R}$ such that, for any finite subset $\{x_i\}_{i=1}^l$ of X , the $l \times l$ matrix $(K(x_i, x_j))_{i,j=1}^l$ is positive semidefinite. The reproducing kernel Hilbert space \mathcal{H}_K associated with a Mercer kernel K is the Hilbert space completed by the span of $\{K_x = K(\cdot, x) : x \in X\}$ under the norm $\|\cdot\|_K$ induced by the inner product $\langle \cdot, \cdot \rangle_K$ satisfying $\langle K_x, K_{x'} \rangle_K = K(x, x')$. The reproducing property in \mathcal{H}_K takes the form $f(x) = \langle f, K_x \rangle_K$ for all $x \in X, f \in \mathcal{H}_K$, which indicates that

$$\|f\|_\infty \leq \kappa \|f\|_K, \quad \forall f \in \mathcal{H}_K, \quad (4)$$

where $\kappa = \sqrt{\sup_{x \in X} |K(x, x)|}$.

Let K be a Mercer kernel. The integral operator $L_K : \mathcal{H}_K \rightarrow \mathcal{H}_K$ given by

$$L_K f = \int_X \int_X f(x) (K_x - K_{x'}) d\rho_X(x) d\rho_X(x') \quad (5)$$

is introduced in [12] to analyze the following regularized ranking algorithm:

$$f_\gamma^z = \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{m^2} \sum_{i,j=1}^m (y_i - y_j - (f(x_i) - f(x_j)))^2 + \gamma \|f\|_K^2 \right\}. \quad (6)$$

This operator is compact, positive, and self-adjoint. In particular, it has at most countably many nonzero eigenvalues

and all of these eigenvalues are nonnegative. Let us arrange these eigenvalues $\{\lambda_i\}$ (with multiplicities) as a nonincreasing sequence tending to 0 and take an associated sequence of eigenfunctions $\{\phi_i\}$ to be an orthonormal basis of \mathcal{H}_K . In the remainder of this paper, we will use a general assumption that

$$f_\rho = L_K^r(g_\rho) \quad \text{for some } r > 0 \text{ and } g_\rho \in \mathcal{H}_K, \quad (7)$$

where the power L_K^r of L_K is defined in terms of $\{\lambda_i\}$ and $\{\phi_i\}$ by

$$L_K^r \left(\sum_{i=1}^{\infty} c_i \phi_i \right) = \sum_{i=1}^{\infty} \lambda_i^r c_i \phi_i. \quad (8)$$

The assumption (7) is equivalent to $f_\rho = \sum_{i=1}^{\infty} \lambda_i^r d_i \phi_i$, where $\{d_i\}$ are the Fourier coefficients of g_ρ with respect to $\{\phi_i\}$; that is, $g_\rho = \sum_{i=1}^{\infty} d_i \phi_i$. Note from (8) that the exponent r measures the decay of the coefficients $\{\lambda_i^r d_i\}$ of f_ρ with respect to the orthonormal basis $\{\phi_i\}$ of \mathcal{H}_K . Thus it can be regarded as a measurement for the regularity of the regression function f_ρ .

Let $\mathbf{x} = \{x_i\}_{i=1}^m$ be the unlabeled part of the samples $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^m$. An empirical version of the operator L_K with respect to \mathbf{x} is given by

$$L_K^{\mathbf{x}}(f) = \frac{1}{m(m-1)} \sum_{i,j=1}^m f(x_i) (K_{x_i} - K_{x_j}), \quad f \in \mathcal{H}_K. \quad (9)$$

The operator $L_K^{\mathbf{x}} : \mathcal{H}_K \rightarrow \mathcal{H}_K$ is self-adjoint and positive with rank at most m . We denote its eigensystem to be $\{(\lambda_i^{\mathbf{x}}, \phi_i^{\mathbf{x}})\}$, where the eigenvalues $\{\lambda_i^{\mathbf{x}}\}$ are arranged in nonincreasing order with $\lambda_i^{\mathbf{x}} = 0$ whenever $i > m$ and the corresponding eigenfunctions $\{\phi_i^{\mathbf{x}}\}$ form an orthonormal basis of \mathcal{H}_K . It can be proved that $\mathbb{E}_{\mathbf{x}}(L_K^{\mathbf{x}}) = L_K$, which means that the eigenfunctions $\{\phi_i\}$ can be approximated by the empirical eigenfunctions $\{\phi_i^{\mathbf{x}}\}$. This fact indicates that the first m eigenfunctions are reasonably promising for ranking.

1.3. The Computation of Empirical Eigenpair. Before proceeding further, we need to show how the empirical eigenpairs $\{(\lambda_i^{\mathbf{x}}, \phi_i^{\mathbf{x}})\}$ can be found explicitly. The main difficulty here is that the kernel of L_K is not symmetric even though it is a self-adjoint operator on \mathcal{H}_K , which makes the computations of empirical eigenfunctions relatively difficult (We refer the readers to [13] for some results on regression learning with indefinite kernels.)

Denote the symmetric matrix $(K(x_i, x_j))_{i,j=1}^m$ by \mathbf{K} and define $\mathbf{A} = (1/m)(m\mathbf{I} - \mathbf{1}\mathbf{1}^T)\mathbf{K}$ and $\mathbf{B} = (1/m)\mathbf{A}(m\mathbf{I} - \mathbf{1}\mathbf{1}^T)$, where \mathbf{I} is the m th-order unit matrix and $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^m$. The proofs of Lemmas 1–5 can be found in [14].

Lemma 1. *Let (λ, v) be an eigenpair of \mathbf{B} . If $\lambda \neq 0$; then $(\lambda, (m\mathbf{I} - \mathbf{1}\mathbf{1}^T)v)$ is an eigenpair of \mathbf{A} .*

It is easy to see that \mathbf{B} is a positive semidefinite matrix. Denote its eigenvalues as $\tilde{\lambda}_1^{\mathbf{x}} \geq \dots \geq \tilde{\lambda}_{t^{\mathbf{x}}}^{\mathbf{x}} > \tilde{\lambda}_{t^{\mathbf{x}}+1}^{\mathbf{x}} = \dots = \tilde{\lambda}_m^{\mathbf{x}} = 0$ with $t^{\mathbf{x}}$ being the rank of \mathbf{B} and the corresponding orthonormal eigenvectors as $\tilde{v}_1, \dots, \tilde{v}_{t^{\mathbf{x}}}, \tilde{v}_{t^{\mathbf{x}}+1}, \dots, \tilde{v}_m$.

Lemma 2. Let $\{(\tilde{\lambda}_l^x, \tilde{v}_l)\}_{l=1}^m$ be eigenpairs of \mathbf{B} . Then, for $1 \leq l \leq m$, one has

$$\begin{aligned} L_K^x \left(\sum_{j=1}^m ((m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \tilde{v}_l)_j K_{x_j} \right) \\ = \frac{\tilde{\lambda}_l^x}{m-1} \left(\sum_{j=1}^m ((m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \tilde{v}_l)_j K_{x_j} \right). \end{aligned} \quad (10)$$

Lemma 3. Let $\{(\tilde{\lambda}_l^x, \tilde{v}_l)\}_{l=1}^m$ be eigenpairs of \mathbf{B} . Then, for $1 \leq p, q \leq m$, one has

$$\begin{aligned} \left\langle \sum_{j=1}^m ((m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \tilde{v}_p)_j K_{x_j}, \sum_{j=1}^m ((m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \tilde{v}_q)_j K_{x_j} \right\rangle_K \\ = m^2 \delta_{p,q} \tilde{\lambda}_q^x. \end{aligned} \quad (11)$$

It follows from Lemmas 2 and 3 that the numbers $\{\tilde{\lambda}_l^x/(m-1)\}_{l=1}^m$ are eigenvalues of the operator L_K^x and the functions $\{(m\sqrt{\tilde{\lambda}_l^x})^{-1} \sum_{j=1}^m ((m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \tilde{v}_l)_j K_{x_j}\}_{l=1}^m$ are the corresponding orthonormal eigenfunctions. Moreover, we have $d^x \geq t^x$, where d^x denotes the rank of L_K^x .

Lemma 4. Let $\{(\lambda_l^x, \phi_l^x)\}_{l=1}^m$ be eigenpairs of L_K^x . Then, for $1 \leq l \leq m$, one has

$$\mathbf{B} (m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \phi_l^x|_{\mathbf{x}} = (m-1) \lambda_l^x (m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \phi_l^x|_{\mathbf{x}}, \quad (12)$$

where $\phi_l^x|_{\mathbf{x}}$ is the vector in \mathbb{R}^m obtained by restricting the function ϕ_l^x onto the sampling points.

Lemma 5. Let $\{(\lambda_l^x, \phi_l^x)\}_{l=1}^m$ be eigenpairs of L_K^x . Then, for $1 \leq p, q \leq m$, one has

$$\left\langle (m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \phi_p^x|_{\mathbf{x}}, (m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \phi_q^x|_{\mathbf{x}} \right\rangle_{\mathcal{L}^2} = m^2 (m-1) \delta_{p,q} \lambda_p^x. \quad (13)$$

By using Lemmas 4 and 5, we come to a conclusion that the numbers $\{(m-1)\lambda_l^x\}_{l=1}^{d^x}$ are eigenvalues of the matrix \mathbf{B} , the vectors $\{(m\sqrt{(m-1)\lambda_p^x})^{-1} (m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \phi_p^x|_{\mathbf{x}}\}_{l=1}^{d^x}$ are corresponding orthonormal eigenvectors, and $d^x \leq t^x$.

Based on the above arguments, we are now confident that the following theorem can be proved, which yields a method for computing the empirical eigenpairs $\{(\lambda_l^x, \phi_l^x)\}_{l=1}^{d^x}$ explicitly.

Theorem 6. The number of positive eigenvalues of L_K^x is equal to that of \mathbf{B} . Moreover, the empirical eigenpairs $\{(\lambda_l^x, \phi_l^x)\}_{l=1}^{d^x}$ of L_K^x can be computed by using the eigenpairs $\{(\tilde{\lambda}_l^x, \tilde{v}_l)\}_{l=1}^{d^x}$ of \mathbf{B} as follows:

$$\lambda_l^x = \frac{\tilde{\lambda}_l^x}{m-1}, \quad (14)$$

$$\phi_l^x = \left(m\sqrt{\tilde{\lambda}_l^x} \right)^{-1} \sum_{j=1}^m ((m\mathbf{I} - \mathbf{1}\mathbf{1}^T) \tilde{v}_l)_j K_{x_j},$$

for $l = 1, \dots, d^x$ with d^x denoting the rank of L_K^x .

1.4. The Ranking Algorithm. Prompted by the above analysis, we propose a learning algorithm for ranking as follows. Let ϵ be a positive number (called a truncated parameter) and let $\mathcal{D}_{m,\epsilon}$ denote the set of empirical eigenfunctions such that their corresponding eigenvalues are less than or equal to ϵ ; that is, $\mathcal{D}_{m,\epsilon} = \{\phi_i^x : 1 \leq i \leq m, \lambda_i^x \geq \epsilon\}$. Let s be the number of eigenfunctions in $\mathcal{D}_{m,\epsilon}$. Our ranking algorithm now takes the form

$$\begin{aligned} c_\epsilon^z = \arg \min_{c \in \mathbb{R}^s} \left\{ \frac{1}{m^2} \sum_{i,j=1}^m \left(y_i - y_j \right. \right. \\ \left. \left. - \left(\sum_{l=1}^s c_l \phi_l^x(x_i) - \sum_{l=1}^s c_l \phi_l^x(x_j) \right) \right)^2 \right\}, \end{aligned} \quad (15)$$

and the output function is

$$f_\epsilon^z = \sum_{l=1}^s c_{\epsilon,l}^z \phi_l^x. \quad (16)$$

We are concerned in this paper with the representer theorem, that is, the explicit solution to problem (15), and the convergence analysis in the \mathcal{H}_K -norm of the above algorithm. Previous work on error analysis of ranking algorithms, such as [3, 8], deals only with generalization properties of the algorithms. Though convergence analysis of classification and regression algorithms has been well studied (see, e.g., [15, 16]), little research has been conducted in establishing similar results in the setting of ranking. Perhaps the first work is that of Chen [12], who derives the convergence rate of a regularized ranking algorithm by means of the technique of operator approximation. Our results can be considered as another attempt in this direction. It should be pointed out that, for the sake of simplicity, rather than taking all target functions into consideration, we will here restrict ourselves to the regression function f_ρ . In other words, we will consider the convergence bounds for $\|f_\epsilon^z - f_\rho\|_K$, instead of $\inf_{f \in \mathcal{F}} \|f_\epsilon^z - f\|_K$.

Notice that, compared with classification or regression problems, the main difference in the formulation of ranking problems is that its performance or loss is measured on pairs of examples, rather than on individual examples. This results in the double-index summation in algorithm (15), which prevents us from directly applying the standard Hoeffding inequality used to obtain convergence bounds for classification and regression. We will tackle this problem by a McDiarmid-Bernstein type probability inequality for vector-valued random variables [17] as is done in [8, 12]. Finally, we show that when the eigenvalues decay polynomially, the algorithm produces sparse representations with respect to the empirical eigenfunctions by choosing a suitable parameter ϵ .

2. The Representer Theorem

In this section we provide the representer theorem for algorithm (15). The key point in proving the representer

theorem is an equality involving the empirical eigenfunctions and eigenvalues.

Theorem 7. *The solution to problem (15) is given by*

$$c_{\epsilon,l}^z = \frac{S_l^z}{2m(m-1)\lambda_l^x}, \quad 1 \leq l \leq s, \quad (17)$$

where $S_l^z = \sum_{i,j=1}^m (y_i - y_j)(\phi_l^x(x_i) - \phi_l^x(x_j))$.

Proof. The empirical error part takes the form

$$\begin{aligned} & \frac{1}{m^2} \sum_{i,j=1}^m \left(y_i - y_j - \left(\sum_{l=1}^s c_l \phi_l^x(x_i) - \sum_{l=1}^s c_l \phi_l^x(x_j) \right) \right)^2 \\ &= \frac{1}{m^2} \sum_{i,j=1}^m \left(\sum_{l=1}^s c_l \phi_l^x(x_i) - \sum_{l=1}^s c_l \phi_l^x(x_j) \right)^2 \\ & \quad - \frac{2}{m^2} \sum_{l=1}^s c_l S_l^z + \frac{1}{m^2} \sum_{i,j=1}^m (y_i - y_j)^2 \\ &= \frac{2}{m^2} \sum_{p,q=1}^s c_p c_q \sum_{i,j=1}^m \phi_p^x(x_i) (\phi_q^x(x_i) - \phi_q^x(x_j)) \\ & \quad - \frac{2}{m^2} \sum_{l=1}^s c_l S_l^z + \frac{1}{m^2} \sum_{i,j=1}^m (y_i - y_j)^2. \end{aligned} \quad (18)$$

A routine computation gives rise to

$$\sum_{i,j=1}^m \phi_p^x(x_i) (\phi_q^x(x_i) - \phi_q^x(x_j)) = m(m-1) \delta_{p,q} \lambda_p^x. \quad (19)$$

By using (19), one can carry on with the above equality chain as follows:

$$\begin{aligned} & \frac{1}{m^2} \sum_{i,j=1}^m \left(y_i - y_j - \left(\sum_{l=1}^s c_l \phi_l^x(x_i) - \sum_{l=1}^s c_l \phi_l^x(x_j) \right) \right)^2 \\ &= \frac{2(m-1)}{m} \sum_{p,q=1}^s c_p c_q \delta_{p,q} \lambda_p^x \\ & \quad - \frac{2}{m^2} \sum_{l=1}^s c_l S_l^z + \frac{1}{m^2} \sum_{i,j=1}^m (y_i - y_j)^2 \\ &= \frac{2(m-1)}{m} \sum_{l=1}^s c_l^2 \lambda_l^x - \frac{2}{m^2} \sum_{l=1}^s c_l S_l^z \\ & \quad + \frac{1}{m^2} \sum_{i,j=1}^m (y_i - y_j)^2. \end{aligned} \quad (20)$$

Hence we have an equivalent form of (15) as

$$c_\epsilon^z = \arg \min_{c \in \mathbb{R}^s} \sum_{l=1}^s \left\{ c_l^2 - \frac{1}{m(m-1)} \frac{S_l^z}{\lambda_l^x} c_l \right\}. \quad (21)$$

The component $c_{\epsilon,l}^z$ can be found by solving the following optimization problem:

$$c_{\epsilon,l}^z = \arg \min_{c \in \mathbb{R}} \left\{ c_l^2 - \frac{1}{m(m-1)} \frac{S_l^z}{\lambda_l^x} c_l \right\}, \quad (22)$$

which has the solution given by (17). This proves the theorem. \square

3. The Error Analysis

In order to derive an error bound for algorithm (15), we need some preliminary inequalities. The following Hoffman-Wielandt inequality establishes the relationship between $\lambda_i - \lambda_i^x$ and $L_K - L_K^x$, which has been investigated in [18–21].

Lemma 8. *One has*

$$\sum_{l=1}^{\infty} (\lambda_l - \lambda_l^x)^2 \leq \|L_K - L_K^x\|_{\text{HS}}^2, \quad (23)$$

where $\|L_K - L_K^x\|_{\text{HS}}$ is the Hilbert-Schmidt norm of $\text{HS}(\mathcal{H}_K)$, the Hilbert space of all Hilbert-Schmidt operators on \mathcal{H}_K , with inner product $\langle A, B \rangle_{\text{HS}} = \text{Tr}(B^T A)$. Here Tr denotes the trace of a linear operator.

The inner product in $\text{HS}(\mathcal{H}_K)$ can also be defined by $\langle A, B \rangle_{\text{HS}} = \sum_j \langle A e_j, B e_j \rangle_K$, where $\{e_j\}$ is an orthonormal basis of \mathcal{H}_K . The space $\text{HS}(\mathcal{H}_K)$ is a subspace of the space of bounded linear operators on \mathcal{H}_K , denoted as $(\mathcal{H}_K, \|\cdot\|)$, with the norm relations $\|A\| \leq \|A\|_{\text{HS}}$ and $\|AB\|_{\text{HS}} \leq \|A\|_{\text{HS}} \|B\|$.

To bound the quantity $\|L_K - L_K^x\|_{\text{HS}}$, we introduce the following McDiarmid-Bernstein type of probability inequality for vector-valued random variable established in [17].

Lemma 9. *Let $\mathbf{z} = \{z_i\}_{i=1}^m$ be independently drawn according to a probability distribution ρ on Z , $(H, \|\cdot\|)$ a Hilbert space, and $F : Z^m \rightarrow H$ measurable. If there is $\bar{M} > 0$ such that $\|F(\mathbf{z}) - \mathbb{E}_{z_i}(F(\mathbf{z}))\| \leq \bar{M}$ for each $1 \leq i \leq m$ and almost every $\mathbf{z} \in Z^m$, then for every $\epsilon > 0$,*

$$\text{Prob} \{ \|F(\mathbf{z}) - \mathbb{E}_{\mathbf{z}}(F(\mathbf{z}))\| > \epsilon \} \leq 2 \exp \left\{ -\frac{\epsilon^2}{2(\bar{M}\epsilon + \sigma^2)} \right\}, \quad (24)$$

where $\sigma^2 = \sum_{i=1}^m \sup_{\mathbf{z} \setminus \{z_i\} \in Z^{m-1}} \mathbb{E}_{z_i} \{ \|F(\mathbf{z}) - \mathbb{E}_{z_i}(F(\mathbf{z}))\|^2 \}$. For any $0 < \delta < 1$, with confidence $1 - \delta$, there holds

$$\|F(\mathbf{z}) - \mathbb{E}_{\mathbf{z}}(F(\mathbf{z}))\| \leq 2(\bar{M} + \sqrt{\sigma^2}) \log \frac{2}{\delta}. \quad (25)$$

From the fact $\text{Tr}(\langle \cdot, K_x \rangle_K K_{x'}) = K(x, x')$ (see [22]) and after tedious calculations, one can derive that, for each $l \in \{1, \dots, m\}$, $\|L_K^x - \mathbb{E}_{x_l}(L_K^x)\|_{\text{HS}} \leq (6/m)\kappa^2$. By Lemma 9, we obtain the following.

Lemma 10. *For any $0 < \delta < 1$, with confidence $1 - \delta$, there holds*

$$\|L_K - L_K^x\|_{\text{HS}} \leq \frac{24\kappa^2 \log(2/\delta)}{\sqrt{m}}. \quad (26)$$

Define a vector-valued function $F : Z^m \rightarrow \mathcal{H}_K$ by

$$F(\mathbf{z}) = \frac{1}{m(m-1)} \sum_{i,j=1}^m (y_i - y_j - (f_\rho(x_i) - f_\rho(x_j))) K_{x_i}. \quad (27)$$

It is easy to show that $\mathbb{E}_{\mathbf{z}}(F(\mathbf{z})) = 0$ and for each $l \in \{1, \dots, m\}$, $\|F(\mathbf{z}) - \mathbb{E}_{\mathbf{z}}(F(\mathbf{z}))\|_K \leq 8M\kappa/m$. By Lemma 9, one has the following.

Lemma 11. *For any $0 < \delta < 1$, with confidence $1 - \delta$, there holds*

$$\left\| \sum_{i,j=1}^m (y_i - y_j - (f_\rho(x_i) - f_\rho(x_j))) K_{x_i} \right\|_K \leq 32\sqrt{m}(m-1)M\kappa \log \frac{2}{\delta}. \quad (28)$$

With the help of the preceding five lemmas, we are now in a position to derive an error estimate for the algorithm. We will conduct analysis for the error in the \mathcal{H}_K -metric, which makes the corresponding error estimate stronger than that performed in the $L^2_{\rho_X}$ -metric [16].

Theorem 12. *Assume (7). For any $0 < \delta < 1$, with confidence $1 - \delta$, one has*

$$\begin{aligned} \|f_\epsilon^{\mathbf{z}} - f_\rho\|_K &\leq \|g_\rho\|_K \lambda_s^r + \frac{C_1 \log(4/\delta)}{\epsilon \sqrt{m}} \\ &\quad + \lambda_s^{\min\{1,r-1\}} \frac{C_3 \log(4/\delta)}{\sqrt{m}} \\ &\quad + C_2 \lambda_s^{\min\{1,r-1\}} \left(\sum_{l=s+1}^{\infty} \lambda_l^{\max\{2r,2\}} \right)^{1/2}, \end{aligned} \quad (29)$$

where C_1, C_2 , and C_3 are constants independent of m and δ (given explicitly in the proof).

Proof. By Lemmas 10 and 11, we know that for any $0 < \delta < 1/2$ there exists a subset Z_δ of Z^m of measure at least $1 - 2\delta$ such that both (26) and (28) hold for each $\mathbf{z} \in Z_\delta$.

Let $\mathbf{z} \in Z_\delta$. It follows from the orthogonal expansion in terms of the orthonormal basis $\{\phi_l^{\mathbf{x}}\}$ that

$$\begin{aligned} \|f_\epsilon^{\mathbf{z}} - f_\rho\|_K^2 &= \sum_{l=s+1}^{\infty} \langle f_\rho, \phi_l^{\mathbf{x}} \rangle_K^2 + \sum_{l=1}^s (\langle f_\rho, \phi_l^{\mathbf{x}} \rangle_K - c_{\epsilon,l}^{\mathbf{z}})^2 \\ &=: \Delta_1 + \Delta_2. \end{aligned} \quad (30)$$

We bound the first term Δ_1 on the right-hand side of (30) by decomposing it further into two parts with $f_\rho = \sum_{j=1}^s \lambda_j^r d_j \phi_j + \sum_{j=s+1}^{\infty} \lambda_j^r d_j \phi_j$:

$$\begin{aligned} \sqrt{\Delta_1} &\leq \left(\sum_{l=s+1}^{\infty} \left\langle \sum_{j=1}^s \lambda_j^r d_j \phi_j, \phi_l^{\mathbf{x}} \right\rangle_K^2 \right)^{1/2} \\ &\quad + \left(\sum_{l=s+1}^{\infty} \left\langle \sum_{j=s+1}^{\infty} \lambda_j^r d_j \phi_j, \phi_l^{\mathbf{x}} \right\rangle_K^2 \right)^{1/2}. \end{aligned} \quad (31)$$

The part with $\sum_{j=s+1}^{\infty}$ is easy to deal with since $\{\phi_l^{\mathbf{x}}\}$ is an orthonormal basis; we have

$$\begin{aligned} &\left(\sum_{l=s+1}^{\infty} \left\langle \sum_{j=s+1}^{\infty} \lambda_j^r d_j \phi_j, \phi_l^{\mathbf{x}} \right\rangle_K^2 \right)^{1/2} \\ &= \left\| \sum_{j=s+1}^{\infty} \lambda_j^r d_j \phi_j \right\|_K \leq \|g_\rho\|_K \lambda_s^r, \end{aligned} \quad (32)$$

where the last inequality follows from $\|d_j\|_2 = \|g_\rho\|_K$. The part with $\sum_{j=1}^s$ can be estimated by the Schwarz inequality as

$$\begin{aligned} &\left(\sum_{l=s+1}^{\infty} \left\langle \sum_{j=1}^s \lambda_j^r d_j \phi_j, \phi_l^{\mathbf{x}} \right\rangle_K^2 \right)^{1/2} \\ &\leq \left(\sum_{l=s+1}^{\infty} \|\{d_j\}\|_2^2 \sum_{j=1}^s \lambda_j^{2r} \langle \phi_j, \phi_l^{\mathbf{x}} \rangle_K^2 \right)^{1/2} \\ &\leq \|g_\rho\|_K \left(\sum_{l=s+1}^{\infty} \sum_{j=1}^s \lambda_j^{2r} \langle \phi_j, \phi_l^{\mathbf{x}} \rangle_K^2 \right)^{1/2}. \end{aligned} \quad (33)$$

We continue to bound $(\sum_{l=s+1}^{\infty} \sum_{j=1}^s \lambda_j^{2r} \langle \phi_j, \phi_l^{\mathbf{x}} \rangle_K^2)^{1/2}$ in two cases.

Case 1 ($r \geq 1$). For $l \geq s+1$, we observe that

$$\begin{aligned} \sum_{j=1}^s \lambda_j^{2r} \langle \phi_j, \phi_l^{\mathbf{x}} \rangle_K^2 &\leq \sum_{j=1}^s 2^{2r-1} (\lambda_l^{2r} + (\lambda_j - \lambda_l)^{2r}) \langle \phi_j, \phi_l^{\mathbf{x}} \rangle_K^2 \\ &\leq 2^{2r-1} \sum_{j=1}^s (\lambda_l^{2r} + \lambda_1^{2r-2} (\lambda_j - \lambda_l)^2) \langle \phi_j, \phi_l^{\mathbf{x}} \rangle_K^2 \\ &\leq 2^{2r-1} \sum_{j=1}^s (\lambda_l^{2r} + 2\lambda_1^{2r-2} (\lambda_j - \lambda_l)^2 \\ &\quad + 2\lambda_1^{2r-2} (\lambda_l^{\mathbf{x}} - \lambda_l)^2) \langle \phi_j, \phi_l^{\mathbf{x}} \rangle_K^2. \end{aligned} \quad (34)$$

From the definition of the Hilbert-Schmidt norm, we have

$$\begin{aligned} \|L_K - L_K^{\mathbf{x}}\|_{\text{HS}}^2 &= \sum_{l=1}^{\infty} \|(L_K - L_K^{\mathbf{x}})\phi_l^{\mathbf{x}}\|_K^2 \\ &= \sum_{j,l=1}^{\infty} (\lambda_j - \lambda_l^{\mathbf{x}})^2 \langle \phi_j, \phi_l^{\mathbf{x}} \rangle_K^2. \end{aligned} \quad (35)$$

By Lemma 8, we get

$$\begin{aligned} \sum_{l=s+1}^{\infty} \sum_{j=1}^s \lambda_j^{2r} \langle \phi_j, \phi_l^{\mathbf{x}} \rangle_K^2 \\ \leq 2^{2r-1} \sum_{l=s+1}^{\infty} \lambda_l^{2r} + 2^{2r+1} \lambda_1^{2r-2} \|L_K - L_K^{\mathbf{x}}\|_{\text{HS}}^2. \end{aligned} \quad (36)$$

Case 2 ($r < 1$). We notice that $\lambda_j^{2r} \leq \lambda_s^{2r-2} \lambda_j^2$ and obtain from the above estimate

$$\begin{aligned} \sum_{l=s+1}^{\infty} \sum_{j=1}^s \lambda_j^{2r} \langle \phi_j, \phi_l^{\mathbf{x}} \rangle_K^2 \\ \leq 2\lambda_s^{2r-2} \sum_{l=s+1}^{\infty} \lambda_l^2 + 8\lambda_s^{2r-2} \|L_K - L_K^{\mathbf{x}}\|_{\text{HS}}^2. \end{aligned} \quad (37)$$

The bounds for the two cases together with (32) give a bound for Δ_1 as

$$\sqrt{\Delta_1} \leq \begin{cases} \|g_\rho\|_K \lambda_s^r + \tau_1, & \text{if } r \geq 1, \\ \|g_\rho\|_K \lambda_s^r + \tau_2, & \text{if } r < 1. \end{cases} \quad (38)$$

Here

$$\begin{aligned} \tau_1 &= 2^r \|g_\rho\|_K \left(\left(\sum_{l=s+1}^{\infty} \lambda_l^{2r} \right)^{1/2} + 2\lambda_1^{r-1} \|L_K - L_K^{\mathbf{x}}\|_{\text{HS}} \right), \\ \tau_2 &= 2 \|g_\rho\|_K \lambda_s^{r-1} \left(\left(\sum_{l=s+1}^{\infty} \lambda_l^2 \right)^{1/2} + 2 \|L_K - L_K^{\mathbf{x}}\|_{\text{HS}} \right). \end{aligned} \quad (39)$$

Now we turn to the second term Δ_2 on the right-hand side of (30). Note that

$$\begin{aligned} \sqrt{\Delta_2} &= \left(\sum_{l=1}^s (\langle f_\rho, \phi_l^{\mathbf{x}} \rangle_K - c_{\epsilon,l}^z)^2 \right)^{1/2} \\ &\leq \frac{1}{\epsilon} \left(\sum_{l=1}^s (\lambda_l^{\mathbf{x}} \langle f_\rho, \phi_l^{\mathbf{x}} \rangle_K - \lambda_l^{\mathbf{x}} c_{\epsilon,l}^z)^2 \right)^{1/2}. \end{aligned} \quad (40)$$

By Theorem 7, we have

$$\begin{aligned} \sum_{l=1}^s (\lambda_l^{\mathbf{x}} \langle f_\rho, \phi_l^{\mathbf{x}} \rangle_K - \lambda_l^{\mathbf{x}} c_{\epsilon,l}^z)^2 \\ = \sum_{l=1}^s \left(\left\langle f_\rho, \frac{1}{m(m-1)} \sum_{i,j=1}^m \phi_l^{\mathbf{x}}(x_i) (K_{x_i} - K_{x_j}) \right\rangle_K \right. \\ \left. - \frac{S_l^z}{2m(m-1)} \right)^2 \\ = \sum_{l=1}^s \left(\frac{1}{m(m-1)} \sum_{i,j=1}^m \phi_l^{\mathbf{x}}(x_i) (f_\rho(x_i) - f_\rho(x_j)) \right. \\ \left. - \frac{1}{m(m-1)} \sum_{i,j=1}^m (y_i - y_j) \phi_l^{\mathbf{x}}(x_i) \right)^2 \\ \leq \left\| \frac{1}{m(m-1)} \sum_{i,j=1}^m (y_i - y_j - (f_\rho(x_i) - f_\rho(x_j))) K_{x_i} \right\|_K^2 \\ \leq \frac{32M\kappa \log(2/\delta)}{\sqrt{m}}. \end{aligned} \quad (41)$$

Thus, for any $0 < \delta < 1$, with confidence $1 - \delta$, we have

$$\sqrt{\Delta_2} \leq \frac{32M\kappa \log(2/\delta)}{\epsilon \sqrt{m}}. \quad (42)$$

Putting the bounds for $\sqrt{\Delta_1}$ and $\sqrt{\Delta_2}$ into (26), we know that, with confidence $1 - \delta$, $\|f_\epsilon^z - f_\rho\|_K$ can be bounded by

$$\frac{32M\kappa \log(2/\delta)}{\epsilon \sqrt{m}} + \|g_\rho\|_K \lambda_s^r + \begin{cases} \tau_1, & \text{if } r \geq 1, \\ \tau_2, & \text{if } r < 1. \end{cases} \quad (43)$$

Let $C_1 = 32M\kappa$, $C_2 = 2^{\max\{r,1\}} \|g_\rho\|_K$, and $C_3 = 24\kappa^2 2^{\max\{r+1,2\}} \lambda_1^{\max\{r-1,0\}} \|g_\rho\|_K$. Then the conclusion of Theorem 12 follows by scaling 2δ to δ . \square

To illustrate the error estimate, we establish learning rate for the algorithm (15) in a special case when the eigenvalues $\{\lambda_l\}$ decay polynomially.

Theorem 13. Assume (7) and for some $1/2r < \alpha_2 \leq \alpha_1 < (1+r)\alpha_2 - (1/2)$ and $D_1, D_2 > 0$, the eigenvalues $\{\lambda_l\}$ decay polynomially as

$$D_1 l^{-\alpha_1} \leq \lambda_l \leq D_2 l^{-\alpha_2}. \quad (44)$$

Let $\epsilon > 0$ such that $s = \lfloor m^{1/(2r\alpha_2-1)} \rfloor$ for $r \geq 1$ or $s = \lfloor m^{1/(2\alpha_2-1)} \rfloor$ for $0 < r < 1$. Then, for any $0 < \delta < 1$, with confidence $1 - \delta$, we have

$$\|f_\epsilon^z - f_\rho\|_K \leq \begin{cases} \bar{C}_1 \log \frac{4}{\delta} m^{-1/2}, & \text{if } r \geq 1, \\ \bar{C}_2 \log \frac{4}{\delta} m^{((\alpha_1(1-r))/(2\alpha_2-1)) - (1/2)}, & \text{if } r < 1, \end{cases} \quad (45)$$

where \bar{C}_1 and \bar{C}_2 are constants independent of m and δ (given explicitly in the proof).

Proof. It follows from the asymptotic behavior of the eigenvalues that

$$\|g_\rho\|_K \lambda_s^r + \frac{C_1 \log(4/\delta)}{\epsilon \sqrt{m}} \leq \|g_\rho\|_K D_2^r s^{-\alpha_2 r} + \frac{C_1 \log(4/\delta)}{\epsilon \sqrt{m}}. \quad (46)$$

For $r \geq 1$, we have

$$\begin{aligned} \sum_{l=s+1}^{\infty} \lambda_l^{\max\{2r,2\}} &= \sum_{l=s+1}^{\infty} \lambda_l^{2r} \leq D_2^{2r} \sum_{l=s+1}^{\infty} l^{-2r\alpha_2} \\ &\leq D_2^{2r} \int_s^{\infty} x^{-2r\alpha_2} dx = \frac{D_2^{2r} s^{1-2r\alpha_2}}{2r\alpha_2 - 1}, \end{aligned} \quad (47)$$

$$\begin{aligned} \lambda_s^{\min\{1,r-1\}} \left\{ C_2 \left(\sum_{l=s+1}^{\infty} \lambda_l^{\max\{2r,2\}} \right)^{1/2} + \frac{C_3 \log(4/\delta)}{\sqrt{m}} \right\} \\ \leq C_2 \lambda_s \frac{D_2^r s^{(1-2r\alpha_2)/2}}{\sqrt{2r\alpha_2 - 1}} + C_3 \lambda_s \frac{\log(4/\delta)}{\sqrt{m}} \\ \leq \frac{C_2 \lambda_1 D_2^r}{\sqrt{2r\alpha_2 - 1}} s^{(1-2r\alpha_2)/2} + C_3 \lambda_1 \frac{\log(4/\delta)}{\epsilon \sqrt{m}}. \end{aligned} \quad (48)$$

Combining (46) and (48), we see that

$$\begin{aligned} \|f_\epsilon^z - f_\rho\|_K &\leq \left(\|g_\rho\|_K D_2^r + \frac{C_2 \lambda_1 D_2^r}{\sqrt{2r\alpha_2 - 1}} \right) s^{(1-2r\alpha_2)/2} \\ &\quad + (C_1 + C_3 \lambda_1) \frac{\log(4/\delta)}{\epsilon \sqrt{m}}. \end{aligned} \quad (49)$$

By choosing $\epsilon > 0$ such that $s = \lfloor \epsilon^{2/(2r\alpha_2-1)} m^{1/(2r\alpha_2-1)} \rfloor = \lfloor m^{1/(2r\alpha_2-1)} \rfloor$, we have

$$\begin{aligned} \|f_\epsilon^z - f_\rho\|_K &\leq \left(\|g_\rho\|_K D_2^r + \frac{C_2 \lambda_1 D_2^r}{\sqrt{2r\alpha_2 - 1}} + C_1 + C_3 \lambda_1 \right) \\ &\quad \times \frac{\log(4/\delta)}{\epsilon \sqrt{m}}. \end{aligned} \quad (50)$$

Similarly, for $0 < r < 1$, we have

$$\begin{aligned} \sum_{l=s+1}^{\infty} \lambda_l^{\max\{2r,2\}} &\leq \frac{D_2^{2r} s^{1-2\alpha_2}}{2\alpha_2 - 1}, \\ \lambda_s^{\min\{1,r-1\}} \left\{ C_2 \left(\sum_{l=s+1}^{\infty} \lambda_l^{\max\{2r,2\}} \right)^{1/2} + \frac{C_3 \log(4/\delta)}{\sqrt{m}} \right\} \\ &\leq \frac{C_2 D_1^{r-1} D_2}{\sqrt{2\alpha_2 - 1}} s^{-\alpha_1(r-1)+((1-2\alpha_2)/2)} \\ &\quad + C_3 s^{-\alpha_1(r-1)} \frac{\log(4/\delta)}{\sqrt{m}}. \end{aligned} \quad (51)$$

Thus, we have, in light of $1/2r < \alpha_2 \leq \alpha_1 < (1+r)\alpha_2 - (1/2)$,

$$\begin{aligned} \|f_\epsilon^z - f_\rho\|_K &\leq \left(\|g_\rho\|_K D_2^r + \frac{C_2 D_1^{r-1} D_2}{\sqrt{2\alpha_2 - 1}} \right) s^{\alpha_1(1-r)+((1-2\alpha_2)/2)} \\ &\quad + (C_1 + C_3) s^{\alpha_1(1-r)} \frac{\log(4/\delta)}{\epsilon \sqrt{m}}. \end{aligned} \quad (52)$$

By choosing $\epsilon > 0$ such that $s = \lfloor \epsilon^{2/(2\alpha_2-1)} m^{1/(2\alpha_2-1)} \rfloor = \lfloor m^{1/(2\alpha_2-1)} \rfloor$, we have

$$\begin{aligned} \|f_\epsilon^z - f_\rho\|_K &\leq \left(\|g_\rho\|_K D_2^r + \frac{C_2 D_1^{r-1} D_2}{\sqrt{2\alpha_2 - 1}} + C_1 + C_3 \right) \\ &\quad \times \left(\log \frac{4}{\delta} \right) \epsilon^{(2\alpha_1(1-r)/(2\alpha_2-1))-1} \\ &\quad \times m^{(\alpha_1(1-r)/(2\alpha_2-1))-(1/2)}. \end{aligned} \quad (53)$$

This completes the proof of the theorem with $\bar{C}_1 = (\|g_\rho\|_K D_2^r + (C_2 \lambda_1 D_2^r / \sqrt{2r\alpha_2 - 1}) + C_1 + C_3 \lambda_1) \epsilon^{-1}$ and $\bar{C}_2 = (\|g_\rho\|_K D_2^r + (C_2 D_1^{r-1} D_2 / \sqrt{2\alpha_2 - 1}) + C_1 + C_3) \epsilon^{(2\alpha_1(1-r)/(2\alpha_2-1))-1}$. \square

Remark 14. The truncated parameter ϵ in algorithm (15) plays a role of the regularization parameter instead of γ . Thus error bounds for our algorithm are closely related to the truncated parameter. Note that our learning rates are given in terms of special choices of the truncated parameter which depends on a priori condition (7). However, methods for determining directly the truncated parameter by the data are more preferable to practical learners. This will be our future research direction.

Remark 15. Note that when r is large enough (meaning that f_ρ has high regularity), the learning rate behaves like $m^{-1/2}$. Moreover, the nonzero coefficients in $f_\epsilon^z = \sum_{l=1}^s c_{\epsilon,l}^z \phi_l^x$ are at most $s = \lfloor m^{1/(2r\alpha_2-1)} \rfloor$, which is much smaller than the sample size m when $r\alpha_2$ is large. Thus, our algorithm produces sparse representations with respect to the empirical eigenfunctions under a mild condition.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Nature Science Foundation of China (nos. 11301052, 11301045, 11271060, 11290143, 61227004, and 61100130), the Fundamental Research Funds for the Central Universities (no. DUT13LK45), and the Fundamental Research of Civil Aircraft (no. MJ-F-2012-04).

References

- [1] S. Clemençon, G. Lugosi, and N. Vayatis, "Ranking and empirical minimization of U -statistics," *The Annals of Statistics*, vol. 36, no. 2, pp. 844–874, 2008.
- [2] S. Cléménçon and N. Vayatis, "Ranking the best instances," *The Journal of Machine Learning Research*, vol. 8, pp. 2671–2699, 2007.
- [3] W. Rejchel, "On ranking and generalization bounds," *Journal of Machine Learning Research*, vol. 13, pp. 1373–1392, 2012.
- [4] A. Slivkins, F. Radlinski, and S. Gollapudi, "Ranked bandits in metric spaces: learning diverse rankings over large document collections," *Journal of Machine Learning Research*, vol. 14, pp. 399–436, 2013.
- [5] G. Blanchard, P. Massart, R. Vert, and L. Zwald, "Kernel projection machine: a new tool for pattern recognition," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '04)*, pp. 1649–1656, 2004.
- [6] H. Chen, H. Xiang, Y. Tang, Z. Yu, and X. Zhang, "Approximation analysis of empirical feature-based learning with truncated sparsity," in *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR '12)*, pp. 118–124, Xi'an, China, July 2012.
- [7] X. Guo and D. X. Zhou, "An empirical feature-based learning algorithm producing sparse approximations," *Applied and Computational Harmonic Analysis*, vol. 32, pp. 389–400, 2012.
- [8] S. Agarwal and P. Niyogi, "Generalization bounds for ranking algorithms via algorithmic stability," *Journal of Machine Learning Research*, vol. 10, pp. 441–474, 2009.
- [9] T. Hu, J. Fan, Q. Wu, and D. X. Zhou, "Learning theory approach to minimum error entropy criterion," *Journal of Machine Learning Research*, vol. 14, pp. 377–397, 2013.
- [10] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.
- [11] F. Cucker and D. Zhou, *Learning Theory: An Approximation Theory Viewpoint*, vol. 24 of *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press, Cambridge, UK, 2007.
- [12] H. Chen, "The convergence rate of a regularized ranking algorithm," *Journal of Approximation Theory*, vol. 164, no. 12, pp. 1513–1519, 2012.
- [13] H. W. Sun and Q. Wu, "Indefinite kernel network with dependent sampling," *Analysis and Applications*, vol. 11, no. 5, Article ID 1350020, 15 pages, 2013.
- [14] M. Xu, Q. Fang, and S. F. Wang, "On empirical eigenfunction-based ranking with l^1 norm regularization," Submitted.
- [15] F. Bauer, S. Pereverzev, and L. Rosasco, "On regularization algorithms in learning theory," *Journal of Complexity*, vol. 23, no. 1, pp. 52–72, 2007.
- [16] S. Smale and D. Zhou, "Learning theory estimates via integral operators and their approximations," *Constructive Approximation*, vol. 26, no. 2, pp. 153–172, 2007.
- [17] S. Mukherjee and D. Zhou, "Learning coordinate covariances via gradients," *Journal of Machine Learning Research*, vol. 7, pp. 519–549, 2006.
- [18] R. Bhatia and L. Elsner, "The Hoffman-Wielandt inequality in infinite dimensions," *Proceedings of the Indian Academy of Sciences: Mathematical Sciences*, vol. 104, no. 3, pp. 483–494, 1994.
- [19] A. J. Hoffman and H. W. Wielandt, "The variation of the spectrum of a normal matrix," *Duke Mathematical Journal*, vol. 20, pp. 37–39, 1953.
- [20] T. Kato, "Variation of discrete spectra," *Communications in Mathematical Physics*, vol. 111, no. 3, pp. 501–504, 1987.
- [21] V. Koltchinskii and E. Giné, "Random matrix approximation of spectra of integral operators," *Bernoulli*, vol. 6, no. 1, pp. 113–167, 2000.
- [22] E. de Vito, A. Caponnetto, and L. Rosasco, "Model selection for regularized least-squares algorithm in learning theory," *Foundations of Computational Mathematics*, vol. 5, no. 1, pp. 59–85, 2005.